NATIONAL AGENDA FOR DIGITAL STEWARDSHIP

2014

NDSA

2014 National Agenda for Digital Stewardship

EXECUTIVE SUMMARY

Effective digital preservation is vital to maintaining the public records necessary for understanding and evaluating government actions, the scientific evidence base for replicating experiments, building on prior knowledge, and the preservation of the nation's cultural heritage. Substantial work is needed to ensure that today's valuable digital content remains accessible, useful, and comprehensible in the future — supporting a thriving economy, a robust democracy, and a rich cultural heritage. The 2014 National Agenda for Digital Stewardship integrates the perspective of dozens of experts and hundreds of institutions, convened through the Library of Congress, to provide funders and other executive decision-makers with insight into emerging technological trends, gaps in digital stewardship capacity, and key areas for development. It is meant to inform, rather than replace, individual organizational efforts, planning, goals, and opinions. Its aim is to offer inspiration and guidance and suggest potential directions and areas of inquiry for research and future work in digital stewardship.

The Agenda outlines the challenges and opportunities related to digital preservation activities in four broad areas: Organizational Roles, Policies, and Practices; Digital Content Areas; Infrastructure Development; and Research Priorities. The sections are arranged from the most comprehensive and encompassing topics to sequentially drill down to more specific challenges and recommendations. The Organizational Roles, Policies, and Practices section discusses the overarching challenges the digital preservation community faces. The Digital Content Areas section highlights specific kinds of content that need attention. The Infrastructure Development section identifies opportunities and makes specific recommendations for how the digital preservation community can respond. The Research Priorities section provides detailed recommendations to prioritize resource allocation towards areas of research that are critical to the advancement of both basic understanding and the effective practice of digital preservation.

Organizational Roles, Policies, and Practices

Despite continued preservation mandates, it has become increasingly difficult to adequately preserve valuable digital content because of a complex set of interrelated societal, technological, financial, and organizational pressures, including:

- Increased scope of responsibilities
  (data management, education of content creators, etc.)
- Growing financial pressures - increased costs and decreasing resources
- Lack of adequate staff, in numbers and expertise
- Increased complexity and volume of data
- Rapidly accelerating technological change
- Evolving data management, security and compliance policies
- Complex and evolving landscape of rights management
- Lack of prioritization of digital preservation by higher administration and those controlling budgets

As a community, we need to dramatically increase cross-organizational cooperation to increase the impact and leverage investments made by individual institutions. We must work together to raise the profile of digital preservation and campaign for more resources and higher priority given to digital preservation, and to highlight the importance of digital curation and the real costs of ensuring long term access. We must also coordinate to develop comprehensive coverage on critical standards bodies, and promote systematic community monitoring of technology changes relevant to digital preservation.

Digital Content Areas

Both born-digital and digitized content present a multitude of challenges to stewards tasked with preservation: the size of data requiring preservation; the selection of content when the totality cannot be preserved; and the selection of modes of both content storage and migration to ensure long-term preservation. Areas of content of particular concern include:

- Electronic Records
- Research Data
- Web and Social Media
- Moving Image and Recorded Sound

Across these areas, content value and selection represent a core challenge that organizations need to address. Furthermore, digital stewardship planning must go beyond a focus on content we already have and technology already in use. Moreover, research is required to develop theoretically grounded and empirically tested models of information valuation.

Technical Infrastructure Development

Successful digital preservation requires taking a broad view of technical infrastructure: the Agenda defines it generally as the set of interconnected technical elements that provides a

framework for supporting an entire structure of design, development, deployment and documentation in service of applications, systems, and tools for digital preservation. This definition includes hardware, software, and systems. Organizational policies, practice, and regulation inform many of the observations and recommendations for the development of digital stewardship infrastructure.

Specific priorities identified for infrastructure investment include:

- File Format Action Plan Development
- Interoperability and Portability in Storage Architectures
- Integration of Digital Forensics Tools
- Ensuring Content Integrity

There is a clear need for organizations to share their assessments of institutional risk and their plans for mitigating those risks; to develop use-case driven best practices for fixity in particular system designs and configurations; and to move the basic research in digital forensics tools from research to implementation in production workflows for organizations. Moreover, the need for integration, interoperability, portability, and related standards and protocols stands out as a theme across all of these areas of infrastructure development.

Research Priorities

Research is critical to the advancement of both basic understanding and the effective practice of digital preservation. Research in digital preservation is under-resourced. In part this is because the payoff from long-term access occurs primarily in the medium-long term and tends to benefit broad and diverse communities.

We expect that research investment in five areas will yield unusually large impact:

- Applied Research for Cost Modeling and Audit Modeling
- Understanding Information Equivalence & Significance
- Policy Research on Trust Frameworks
- Preservation at Scale
- The Evidence Base for Digital Preservation

A common challenge running through this report is the limited amount of empirical evidence available. The digital preservation community is beginning to develop a shared evidence base that can be used to answer these and similar questions. However, these studies must be broadened and repeated over time to establish a robust evidence base from which generalizable guidance can be drawn. Furthermore, decision-makers should recognize that basic research in

these areas often needs to be paired with the development, support, and evaluation of infrastructure.

Conclusion

Effective digital stewardship is vital to maintaining the nation's cultural heritage, scientific evidence base, and the public records necessary for understanding and evaluating government actions. The 2014 National Agenda for Digital Stewardship identifies the key technological trends, gaps in digital stewardship capacity, and opportunities for future work for digital preservation professionals, decision-makers, and others interested in investing in the long-term management of digital content.

2014 National Agenda for Digital Stewardship

INTRODUCTION

Our culture is a digital culture. The photographs of our families; the electronic communities where we share and receive news; the maps that give us new insight on where we're going and how to get there; the films and music that shape our shared experiences—all digital. Our digital creations represent an incalculable investment in time, energy, and resources that require responsible care to remain viable over time.

Digital stewardship is a series of managed activities, policies, strategies, and actions that ensure that digital content of vital importance to the nation is acquired, managed, organized, preserved, and accessible for as long as necessary. Digital stewardship activities protect important content in spite of changes in technology, economic sustainability, or institutional capacity.

The *2014 National Agenda for Digital Stewardship* highlights emerging technological trends, identifies gaps in digital stewardship capacity, and provides funders and decision-makers with insight into the work needed to ensure that today's valuable digital content remains accessible, useful and comprehensible in the future, supporting a thriving economy, a robust democracy, and a rich cultural heritage. It is meant to inform, rather than replace, individual organizational efforts, planning, goals, or opinions. It offers inspiration and guidance and suggests potential directions and areas of inquiry for research and future work in digital stewardship.

The *Agenda* is sponsored by the National Digital Stewardship Alliance. The NDSA is a voluntary membership organization of leading government, academic, and private sector organizations with digital stewardship responsibilities. Members of the NDSA collaborate to establish, maintain, and advance the capacity to preserve our nation's digital resources for the benefit of present and future generations.

The NDSA secures and broadens access to the expanding digital resources of the United States of America, develops and coordinates sustainable infrastructures for the preservation of digital content, advocates standards for the stewardship of digital objects, builds a community of practice, promotes innovation, facilitates cooperation between previously unaligned sectors,

and raises awareness of the enduring value of digital resources and the need for active stewardship.

With its national focus, the NDSA is in a unique position to identify and communicate the challenges, opportunities, and priorities for digital stewardship activity in the United States. The NDSA joint leadership group, digital stewardship experts elected from a cross-section of diverse sectors of the U.S. economy, including libraries and archives, academic, technology and commercial concerns, authored this strategic agenda.

The *Agenda* outlines the challenges and opportunities related to digital preservation activities in four broad areas: Organizational Roles, Policies, and Practices; Digital Content Areas; Infrastructure Development; and Research Priorities. The sections are arranged to go from the most comprehensive and encompassing topics to more specific challenges and recommendations. The Organizational Roles, Policies, and Practices section discusses the overarching challenges the digital preservation community faces. The Digital Content Areas section highlights specific kinds of content that need attention. The Infrastructure Development section identifies opportunities and makes specific recommendations for how the digital preservation community can respond. The Research Priorities section provides detailed recommendations for future efforts in areas of research that are critical to the advancement of both basic understanding and the effective practice of digital preservation.

ORGANIZATIONAL ROLES, POLICIES, AND PRACTICES

Despite continued preservation mandates, it has become increasingly difficult to adequately preserve digital content because of a complex set of interrelated societal, technological, financial and organizational pressures:

- Increased scope of responsibilities
  (data management, education of content creators, etc.)
- Growing financial pressures - increased costs and decreasing resources
- Lack of adequate staff, in numbers and expertise
- Increased complexity and volume of data
- Rapidly accelerating technological change
- Evolving data management, security and compliance policies
- Complex and evolving landscape of rights management
- Lack of prioritization of digital preservation by higher administration and those controlling budgets

The pressures listed above are interrelated; therefore, the most effective solutions will address multiple factors. The whole suite of problems must be addressed together for the most effective change. We discuss these pressures in the remainder of this section and in the following section describing digital content areas.

As a community, we must work together to raise the profile of digital preservation and campaign for more resources and higher priority given to digital preservation. There is a need to increase outreach activities and education about the importance and real cost of digital preservation.

We need to dramatically increase cross-organizational cooperation and division of labor to multiply the breadth of impact and investments made within individual institutions. It remains impractical for every institution to develop expertise in every aspect of the digital preservation challenge; different institutions could specialize in different aspects and rely on each other for some functions. If each institution does not have the resources to fully fund all the digital preservation responsibilities and activities, having each institution spend on something different and sharing capabilities with each other would place investments wisely where they could make a real impact. If each institution cannot hire the required number of staff and variety of types of expertise, collaborative hiring and sharing of staff and skills could help. It is key to identify preservation functions that could be outsourced (the NDSA Digital Preservation Staffing Survey[1] revealed some functions) versus the functions that each organization prefers to or must do for itself (e.g. planning, alignment with parent organization's goals and designated communities). It is also essential to establish a network of preservation service providers who can provide different specialized services so every organization does not have to provide all the services it needs for itself. This requires making visible the different services offered, areas of expertise, and standards activities of organizations active in the digital preservation community. The community could then use that visibility to find opportunities where multiple organizations could benefit from a division of labor and identify gaps where something necessary is not getting done. This work would allow members in the community to identify potential specializations, then publicize commitments of organizations to specialize in a particular function so others can begin to rely on it.

This high level of collaboration between many organizations requires several support elements to be in place.[2] The work that still needs to be done is at a community level and includes

[1] NDSA Digital Preservation Staffing Survey http://www.digitalpreservation.gov/ndsa/documents/NDSA-staff-survey-poster-ipres2012.pdf

[2] See for example *Hess, C. ; Ostrom, E. Book Title: Understanding Knowledge as a Commons:From Theory to Practice Publisher : MIT Press;* Nancy McGovern & Katherine Skinner. 2012, Aligning National Approaches to Digital Preservation. Educopia Institute.

developing mature certification and trust frameworks, building organization's capacity to demonstrate trustworthiness, encouraging wide adoption of interoperability standards that would allow organizations to rely on each other more easily for predictable and equivalent outcomes, and establishing a method of providing assurance that the digital preservation community is participating in all relevant standards bodies so that institutions can trust that their digital preservation interests are being represented by someone in the community when it matters.

The digital preservation community needs comprehensive coverage on all critically relevant standards bodies, and coordination so that it is clear who has taken responsibility for what. The community also needs to collaborate on systematic processes of monitoring and responding to technology, rather than reacting to technology changes.[3] Additionally, there is a need to identify more cost-efficient methods of preservation. Research needs to be conducted on cost-efficient but effective preservation, and sustainable financing/billing models.

The digital preservation community also needs to develop and share digital preservation training and staffing resources, especially for training or hiring digital preservation staff (e.g. curricula, training materials, position descriptions). Digital preservation professionals continue to be needed in the intersection between IT and content creators to ensure the longevity of content.[4] There continues to be a need for professional, trained curatorial staff to perform the functions of digital stewardship activity. As the stewardship of digital materials becomes a responsibility for an increasing number and variety of institutions, education, training, and workforce development are seen as key elements in supporting the expertise necessary for building a qualified base of current and future of digital stewards.[5]

Key issues in this area include exploring more practical, immersive internships and fellowship for undergraduates and graduate students, the need for greater fluency with technologies across the field, more robust and affordable professional development opportunities, the economics and efficacy of online learning, better understanding of career paths and organizational roles for digital curators and preservationists, affiliations with data management and preservation in non-humanities disciplines, and exploring collaborative opportunities between educational programs, students, and employers in the digital preservation community.

[3] See Nancy McGovern 2009, *Technology Responsiveness for Digital Preservation: A Model*. University College of London. PhD. Thesis.

[4] DigCurV Project, Training Opportunities survey and Evaluation framework & Training needs survey http://www.digcur-education.org/eng/Resources

[5] Closing the Digital Curation Gap: A Grounded Framework for Providing Guidance and Education in Digital Curation http://www.ils.unc.edu/callee/p57-tibbo.pdf

Continued exploration of new curriculum models for graduate education as well as innovative training and professional development mechanisms for those currently in the field is needed.[6] Potential ways in which training for digital preservation curatorial roles can be extended to a broader reach of professionals should also be explored.

DIGITAL CONTENT AREAS

Both born-digital and digitized content present a multitude of challenges to stewards tasked with preservation: the size of data requiring preservation, the selection of content when the totality cannot be preserved, and the selection of modes of both content storage and format migration to ensure long-term preservation.

Digital stewardship planning must go beyond a focus on content we already have and technology already in use. Even in the near term, a number of trends are evident. Given the ever growing quantity of digital content being produced, scalability is an immediate concern. More and more people globally have access to tools and technologies to create digital content, increasingly with mobile devices equipped with cameras and apps developed specifically for the generation and dissemination of digital content. Moreover, the web continues to be a publishing mechanism for individuals, organizations, and governments,[7] as publishing tools become easier to use. In light of these trends, the question of how to deal with "big data" is a major concern for digital preservation communities.

Content selection policies vary widely depending on the organization and its mission, and when addressing its collections, each organization must discuss and decide upon approaches to many questions. While selection policies for traditional content are most often topically organized, digital content categories, described here, present specific challenges. In the first place, there is the challenge of countering the public expectation that everything digital can be captured and preserved -- stewards must educate the stakeholders on the necessity of selection. Then there are the general organizational questions that apply to all digital preservation collections. For example, how to determine the long-term value of content?

---

[6] *The NMC Horizon Report: 2012 Higher Education Edition*
http://www.educause.edu/ir/library/pdf/HR2012.pdf

[7] *RMICC - Records Management Interagency Coordinating Council Biennial Report 2011-2012.*The State of Texas, October 2012. See Appendix A, Review of Electronic-Records management Practices at Texas State Agencies and Institutions of Higher Education http://www.rmicc.state.tx.us/docs/rmicc11_12biennialrpt.pdf. From Page 5 of 34 in Appendix A: Overall, our findings show that the volume of electronic records in various applications is expanding rapidly over time, and management of the records lags behind available technology. We find that electronic records management (ERM) including email and social media management, face escalating problems that are not adequately supported by current practices, professional skill sets, placement and strategic planning.

Audiences increasingly desire not only access, but enhanced use options and tools for engaging with digital content. Usability is increasingly a fundamental driver of support for preservation, particularly for ongoing monetary support. Which stakeholders should be involved and represented in these determinations? Of the content that is of interest to stakeholders, what is at risk, and must be preserved? What are appropriate deselection policies? What editions/versions, expressions and manifestations (e.g. items in different formats) should be selected?

Electronic Records

The potential loss of electronic records, and the loss of the underlying information these records contain, poses a significant threat to the American memory.[8] Whether it's an electronic diary, an email exchange, or the documentation of government transactions, each of these records is at risk of disappearing unless thoughtful action is taken to preserve important information. Preserving electronic records efficiently and cost-effectively remains a tremendous challenge that needs to be addressed on many levels.[9] The volume of records generated and held by individuals and institutions in electronic format requires changes to traditional paper-based procedures. Rather than relying on file clerks to organize and store information, the information creator– each institution and individual - will be responsible for properly managing his or her own electronic records.[10] A proper infrastructure, supplemented with public outreach, will be critical in educating the public about current deficiencies in long-term electronic preservation and in equipping them to properly save important materials.

Research Data

Some of the most acute challenges of digital content can be illustrated by considering the curation of digital research data. The sheer scale of research data represents a daunting curatorial task. With newly developed scientific instrumentation and the growing use of computer simulations, a research team can generate many terabytes of data per day. Data

---

[8] *Future Watch: Strategies for Long-Term Preservation of Electronic Records*. Hoke, Gordon EJ, CRM. Information Management Journal 46.3 (May/June 2012): 26-28, 30-31, 47. Through Proquest – http://search.proquest.com/docview/1019286317

[9] One attempt at addressing the problem is the Presidential Directive on Records Management http://www.whitehouse.gov/sites/default/files/omb/memoranda/2012/m-12-18.pdf

[10] "to have an effective records management program, agency records management staff must have a baseline of knowledge about electronic records and how to manage them. Records staff do not need to be technological experts, but they have to understand certain fundamental principles and practices of managing electronic records." http://www.archives.gov/records-mgmt/resources/self-assessment-2011.pdf ; *2011Records Management Self-Assessment Report*, NARA.

curators face management at the petabyte scale (1,000 terabytes) and well beyond. Scientific fields such as particle physics, with its collider data, and astronomy, with its sky surveys, as well as research fields and methods like bioinformatics, crystallography, and engineering design generate massive amounts of digital data. Large-scale digitized content being created by initiatives like the Google Books project pose similar challenges.

Although some research data are no more complex than other objects that are routinely curated, a portion of digital research data are complicated to curate. Research data can be heterogeneous, ranging from numeric and image based, to textual, geospatial and other forms. There are many different information standards used (and not used) as well as many different approaches to information structure (e.g., XML-structured documents vs. fixed image and textual file formats). Moreover, the research communities that produce data are equally diverse; data management practices vary greatly both within and between disciplines. There may also be commercial interests in the data and associated data practices.

Perhaps the overriding challenges in all respects to digital research data are the affiliated costs. Domain researchers, technologists, information scientists, and policymakers are searching for sustainable economic models with the ability to accurately predict costs and to balance them across the lifecycle (e.g. costs for ingest, archival management, and dissemination), and through federated inter-institutional repository systems.

Managing research data will also require stewards to take on new roles. These may include enabling researchers to curate their research, absent professional expertise; to apply and adapt metadata in new ways; and to collaborate with researchers in developing new workflows, models and tools.

There is no one-size-fits-all approach when it comes to resolving the management challenges of research data; however, [11] progress might be made by mobilizing the digital preservation and curation community toward in-depth study of these challenges of scale, complexity, research community practice, and cost with the aim of developing new recommendations and potential long-term solutions.

Web and Social Media

While cultural heritage organizations and others have been preserving web content since 1996, the task of preserving born-digital web content has become progressively more difficult as websites become more complex and the scale of the web continues to grow. Crawlers used to collect content as well as access tools used to render the web archives are increasingly challenged to keep up with the explosion of ever-increasingly complex technologies:

[11] Lyon, Liz. "Dealing with Data: Roles, Rights, Responsibilities and Relationships. Consultancy Report." (2007); Graham Prior, 2012 Managing Research Data, Facet Press.

multimedia, mashups, deep-web, databases, and the growing prevalence of heavily scripted site navigational paradigms that do not prevent the collection of data but make replay nearly impossible without changes to an archive visitor's browser configuration. More and more content created or published on the web cannot be preserved using currently available tools.

The International Internet Preservation Consortium (IIPC)[12] developed the Heritrix web crawler[13] and is working to develop a community to stabilize, improve, and support this open source tool in the future. Broad involvement by web archivists not involved directly in IIPC is required, because the development and exploration of improved access tools, including data mining tools for large datasets of web archives, are also necessary. Full-text indexing of web archives continues to challenge researchers and the web archivist community, particularly as archives expand and reach multiple terabyte and petabyte size.

The increasing use of social media by organizations and individuals can also be a challenge to preserve, as services hosting this content do not take preservation as a business model and the changes they make in how they present content can upset the preservation process. The rate of publication/site implementations and change for large social media aggregation sites is, on average, every 3-6 weeks. This makes it virtually impossible to keep pace with an archival quality capture of these resources, unless granted direct access to the site feeds, which is not available to most cultural heritage institutions even for a fee.

Tools developed in recent years, primarily to meet the needs of business compliance regulations, are able to capture more of this type of material on a small scale. These tools are inadequate to the demands complexity and scale of the modern web. Web archiving technologies need to be translated to open source tools that scale to the needs of cultural heritage institutions and others collecting large amounts of data.

In addition to these concerns, there are privacy issues with web and social media. Both websites and social media rely on the cycles of creation of content and connections among content sources and users. At the same time, there is an interest in capturing and preserving the content creation and connections, including the unfolding of those cycles over the lifetime of the content and the connections. The challenges for an individual desiring to define and sustain a particular level and type of privacy in our connected online worlds have been well-documented. Less studied have been the related privacy challenges that exist for institutions archiving the digital files that result from the capture and preservation of websites and social media interactions. It may be useful to conduct research and analysis on the potential application of existing privacy- and personal-security- related practices from other areas, such as university human research protection programs or organizations such as banks and government agencies, to digital preservation practices.

[12] IIPC netpreserve.org

[13] http://sourceforge.net/projects/archive-crawler/

Moving Image and Recorded Sound

One hundred years ago it was very hard to make a movie, as the process required skills, tools and materials that were difficult to obtain. Even as little as forty years ago, prior to the creation of the Betacam, the technology to create movies was not easily accessible to the public. Audio recording was likewise inaccessible to most individuals. Now, however, many of us carry devices in our pockets that can record, edit and disseminate our audio and moving image recordings. This flood of digital content poses many vexing problems for the stewards of these materials. Digital preservation and stewardship of motion picture film, audio, and video presents a multitude of challenges. There is a need for both new standards and for the evolution of existing standards, such as preservation quality reformatting and a myriad of issues that arise from creating and managing large files—not only storage, but the long-term ability to manage and play back these files. Relatedly, while movie and recording industries *should* collaborate with cultural heritage institutions, this is not always the case. It is vital that both content creators and stewards work together to develop standards and workflows that will ensure long-term access to our recorded and moving image heritage.[14]

TECHNICAL INFRASTRUCTURE DEVELOPMENT

Technical infrastructure can be generally defined as the set of interconnected technical elements that provide a framework for supporting an entire structure of design, development, deployment, and documentation in service of applications, systems, and tools for digital preservation. This includes hardware, software, and systems. Organizational policies, practices, and regulations inform many of the observations and recommendations for the development of digital stewardship technical infrastructure.

File Format Action Plan Development

The sustainability of digital file formats and the risks of file format obsolescence persist as significant challenges for stewardship organizations.[15] Now that stewardship organizations are amassing large collections of digital materials, it is important to shift from more abstract considerations about file format obsolescence to develop actionable strategies for monitoring and mining information about the heterogeneous digital files the organizations are managing.

[14] Significant contribution to this section was made by Jimi Jones, Archivist of the Hampshire College Image, Sound and Text Archives: http://www.hampshire.edu/library/index_archives.htm

[15] Arms, Caroline & Fleischhauer, Carl. Digital Formats: Factors for Sustainability, Functionality, and Quality. IS&T Archiving 2005 Conference, Washington, D.C. http://memory.loc.gov/ammem/techdocs/digform/Formats_IST05_paper.pdf

Recent analysis of image formats and HTML doctypes[16] offers a valuable example of how organizations can analyze and share their data for analysis by third-party digital preservation researchers. By collecting and sharing this kind of data, it becomes possible for stewardship organizations to shift toward the development of file-format action plans based on the size of the risks that particular obsolescence threats pose to the significance those formats play in the organizations' managed content. Implementation of tools and services for creating file-format action plans is needed to make timely execution of file format plans a reality for data stewards.

It is also necessary for organizations to itemize and assess the digital content they are actively managing. Stewardship organizations should prioritize the kinds of the development of file format action plans that most appropriately reflect the kinds of content they are actually managing. The common development of processes for executing format action plans would benefit all organizations within the digital preservation community. The digital preservation community would further benefit from organizations sharing their assessments of institutional risk and their plans for mitigating that risk and addressing file format problems with specific plans.

Migration, Interoperability, and Content Integrity

In 2011, when NDSA members were asked about plans for storage systems and architectures, 64% of respondents agreed or strongly agree that their organization planned to make significant changes in technologies in their preservation storage architecture within three years.[17] This underscores a fact that digital preservation practitioners already know quite well, that digital preservation is made possible through a long chain of migration through layers of current hardware and software systems to yet-to-be-established future infrastructures. This highlights the need for interoperability across different layers in these systems. In addition, easy migration of digital content from one system to another between organizations, such as vendor to client or partner to partner, would benefit the community enormously, particularly fostering the building of coalitions around preservation.

This indicates a clear need for standards and the development of model plans for ensuring end-to-end data integrity in these migrations going forward. Much of the current practice is developed on an ad hoc, one-off basis. Given that the forward cycle of migration will clearly be a continual part of digital preservation work, it is essential to develop clear guidance on how to plan for and manage these changes, and how to measure systematically the quality of the

[16] Jackson, Andy. *Formats over Time: Exploring UK Web History.* http://arxiv.org/pdf/1210.1714v1.pdf 5 Oct 2012.

[17] Micah Altman, Jefferson Bailey, Karen Cariani, Michelle Gallinger, Jane Mandelbaum, Trevor Owens, 2013, NDSA Storage Report: Reflections on National Digital Stewardship Alliance Member Approaches to Preservation Storage Technologies. *D-Lib Magazine* 19(5/6)

results. This kind of guidance development would inevitably point to issues that require further development of protocols and standards for interoperability and evaluation to help ensure continuity.

There is also a need for the development of standards, practices and strategies that directly address migration both lateral migration and forward migration. Case studies and more systematic reviews of activities currently underway need to be shared throughout the digital preservation community.

As stewardship organizations manage increasingly large and complex data sets, the need for interoperability at various levels within the technical hardware and software stacks that make digital preservation becomes increasingly important. This includes interoperability of storage devices, hardware, data tape, and file systems software and would help alleviate bottlenecks in the interrelationship between distinct functions in digital preservation workflows. For example, one NDSA member working with large moving image collections expressed a need, shared by many, to be able to send moving image materials to a vendor to digitize, lay on a LTO tape which could then be directly plugged into their tape library without having to re-ingest.

There is a significant need for establishing and promoting technical means by which lower levels of the technology stack can directly integrate without requiring extensive computation and processing at higher levels. Moreover, it would be valuable if there were standards and protocols for interoperability. The digital preservation community can encourage this level of interoperability by identifying and promoting standards, elements, metadata, and other topics that would increase the possibility for technical interoperability of tape storage.

Fixity checking is of particular concern in ensuring content integrity. Abstract requirements for fixity checking can be useful as principals, but when applied universally can actually be detrimental to some digital preservation system architectures. The digital preservation community needs to establish best practices for fixity strategies for different system configurations. For example, if an organization were keeping multiple copies of material on magnetic tape and wanted to check fixity of content on a monthly basis, they might end up continuously reading their tape and thereby very rapidly push their tape systems to the limit of reads for the lifetime of the medium.

There is a clear need for use-case driven examples of best practices for fixity in particular system designs and configurations established to meet particular preservation requirements. This would likely include description of fixity strategies for all spinning disk systems, largely tape-based systems, as well as hierarchical storage management systems. A chart documenting the benefits of fixity checks for certain kinds of digital preservation activities would bring clarity and offer guidance to the entire community. A document modeled after the NDSA Levels of Digital Preservation would be a particularly useful way to provide guidance and information about fixity checks based on storage systems in use, as well as other preservation choices.

Integration of Digital Forensics Tools

As more digital materials are selected for long-term digital preservation, the need to integrate digital forensics tools into production workflows for collections becomes increasingly important. This will require identifying the boundaries between technical infrastructure development and organizational policies, and where there is tension that creates issues for providing access or pursuing work that reduces tension whether it be new or refined policies or services and tools development. Integration of these tools can build on exploratory work using digital forensics.[18] Tools currently under development[19] can be leveraged and workflows[20]can be implemented. Aside from the need for tools and workflow developments, there are also important opportunities for organizations to share resources in order to tackle these issues. In this respect, pioneering new organizational models for centers of stewardship, such as SWAT sites,[21] can help to support the development of centers of excellence that help to scale up this kind of activity.

There is a clear need to move the basic research in digital forensics tools from research to implementation in production workflows for organizations. This would require investment in scaling up tools and creating collaborative models for sharing resources to make this work possible. The digital preservation community would also benefit from a shared space for exchanging knowledge around how forensics tools are being integrated into production preservation activities.

RESEARCH PRIORITIES

This section focuses on areas of research that are critical to the advancement of both the basic understanding and the effective practice of digital preservation. Research in digital preservation is under-resourced, in part because the payoff from long-term access occurs primarily in the medium-long term, and tends to benefit broad and diverse communities. Funding is needed in this area to develop basic theoretical models, extend the evidence base, and to translate research findings into digital preservation practices and tools. Furthermore, digital preservation research is often closely tied to the development and evaluation of infrastructure, which makes

[18] Matthew G. Kirschenbaum, Richard Ovenden, Gabriela Redwine, Rachel Donahue. 2010, Digital Forensics and Born-Digital Content in Cultural Heritage Collections. Council on Library and Information Resources publication ; no. 149. http://www.clir.org/pubs/abstract/reports/pub149 and Jeremy Leighton John, Digital Forensics and Preservation. 2012, DPC Technology Watch Report. DOI: http://dx.doi.org/10.7207/twr12-03

[19] Such as BitCurator (http://www.bitcurator.net/)

[20] AIMS Work Group. 2012. AIMS Born-Digital Collections: An Inter-Institutional Model for Stewardship.

[20]http://www2.lib.virginia.edu/aims/whitepaper/AIMS_final.pdf

[21] Ricky Erway, 2012. Swatting the Long Tail of Digital Media: A Call for Collaboration. Dublin Ohio: OCLC Research. http://www.oclc.org/research/publications/library/2012/2012-08.pdf

it challenging to fund through basic research funding mechanisms. Decision-makers should recognize that basic research in these areas often needs to be paired with the development, support, and evaluation of infrastructure.

Applied Research for Cost Modeling and Audit Modeling

In the near term, there are specific areas of applied research around digital preservation lifecycle issues that need attention. Currently there are limited models for cost estimation for ongoing storage of digital content; cost estimation models need to be robust and flexible. Furthermore, as discussed below in *Strengthening the Evidence Base for Digital Preservation,* there are virtually no models available to systematically and reliably predict the future value of preserved content.

Different approaches to cost estimation should be explored and compared to existing models with emphasis on reproducibility of results. The development of a cost calculator would benefit organizations in making estimates of the long-term storage costs for their digital content. Auditing models also need to be strengthened and further developed, as discussed below in the section on *Policy Research in Trust Frameworks*. The SafeArchive[22] system and other bit-level auditing practices could be connected to the NDSA Levels of Preservation[23] work to help organizations determine and validate the costs of scaling different auditing schemes.

This research needs to address multiple storage models: Locally stored data, distributed preservation networks, data cooperatives, cloud storage, brokered cloud storage systems and hybrid systems should each be addressed in cost models and auditing practices so that organizations can make informed cost-effective digital preservation decisions.

Understanding Information Equivalence and Significance

The multiplicity of instantiations of the same or similar digital objects illustrates the need for basic research and the application of basic research to explore the many ways multiple digital objects could contain equitant informational content given different contexts of significance. For instance, a single photograph may be represented by any number of derivative files of varying sizes, in varying formats, and with different sets of embedded metadata inside it.[24] Similarly, an

[22] www.safearchive.org

[23] http://blogs.loc.gov/digitalpreservation/2012/11/ndsa-levels-of-digital-preservation-release-candidate-one/

[24] C. Marshall Digital Copies And A Distributed Notion Of Reference In Personal Archives In *Digital Media: Technological And Social Challenges Of The Interactive World* edited by W. Aspray, M Winget.

organization may have 15 PDFs of the same article each with a different cover page, but all of which are substantively identical. Preservation research needs to map out the networks of similarity and equivalence across different instantiations of objects so that they can make better decisions on how to manage content, bearing in mind what properties of a given set of digital objects are significant[25] to their particular community of use. Research is also required in order to characterize quality and fidelity dimensions and create methods for computing format-independent fingerprints of content,[26] so that the fidelity of digital objects can be effectively managed over time. In this space, there is potential value in fuzzy hashing algorithms that can map out the similarity of bitstreams, applications to analyze and compare rendered content in different formats (image comparison, extracting and comparing sound frequencies across audio and video files, etc.), and other innovative potential modes for asserting that some aspect of a given set of objects is similar in a particular way to another set of objects. Beyond basic research to develop methods for identifying information equivalence, there is a need for research in different usage contexts to understand when particular modes or levels of information equivalence are relevant to particular stakeholders in particular contexts.

Policy Research on Trust Frameworks

There is a well-identified taxonomy of potential single-points-of-failure (highly correlated risks), that at minimum, a trustworthy preservation system should mitigate: These risks include media failure, hardware failure, software failure, communication errors, network failure, media and hardware obsolescence, software obsolescence, operator error, natural disaster, external attack, internal attack, economic failure, and organizational failure.[27]

Geographic risk, curatorial error, internal malfeasance, economic failure, and organizational failure require that replications be diversified across distributed, and often, collaborative organizations.[28] No one provider can or should provide all elements of long-term preservation --

[25] Hedstrom, Margaret, and Christopher A. Lee. "Significant properties of digital objects: definitions, applications, implications." *Proceedings of the DLM-Forum*. 2002.

[26] Altman, Micah. "A fingerprint method for scientific data verification." *Advances in Computer and Information Sciences and Engineering*. Springer Netherlands, 2008. 311-316.

[27] D. S. H. Rosenthal, T. Robertson, T. Lipkis, V. Reich, and S. Morabito, "Requirements for Digital Preservation Systems", D-Lib Magazine 11 (11), 2005 http://www.dlib.org/dlib/november05/rosenthal/11rosenthal.html

[28] Altman, M., Beecher, B., & Crabtree, J. (2009). A Prototype Platform for Policy-Based Archival Replication. Against the Grain, 21(2), 44-47. http://www.box.net/shared/gxdcnsxunlpg9xol5h1t
Altman, M., & Crabtree, J. (2011). Using the SafeArchive System : TRAC-Based Auditing of LOCKSS. Archiving 2011 (pp. 165-170). Society for Imaging Science and Technology. http://www.box.net/shared/8py6vl9kxivo6u21rkn8

therefore developing approaches for collaborative stewardship and for modularized review, auditing and certification is required.

In this area, community use of collaborative institutional mechanisms to mitigate preservation risk is growing. This is reflected in the growth of organizations such as the Global LOCKSS Network, Data-PASS, MetaArchive, the Digital Preservation Federation, and the Digital Preservation Network. These organizations, and the multi-institutional stewardship approach they represent, have increased both in use and in recognition.

Nonetheless, the reliability, design, and behavior of both centralized and distributed preservation networks is just beginning to be understood. It is critical to develop robust trust frameworks that address these risks, because institutions need to be able to measure and evaluate and monitor the reliability and trustworthiness of trustworthy repositories, collaborating organizations and third-party services (such as cloud computing). Measuring and evaluating the trustworthiness of such organizations and services is a substantial challenge for policy research.

The preservation community has made progress in this direction: Many of the processes identified with trustworthy content stewardship have been recognized, standardized, and documented in the Trustworthy Repositories Audit & Certification (TRAC) criteria and in the newly released ISO standard that has succeeded it. Furthermore a number of organizations are moving toward formal certification under this ISO 16363 standard.

Notwithstanding, much remains to be done. Few stewardship organizations have obtained trustworthy certification, and a relatively small percentage of stewardship organizations are seeking it,[29] while many third-party services will not seek it at all. Furthermore, no certification process has yet been widely recognized by the preservation community.

A number of approaches to certification auditing and assessment are particularly promising. Self-evaluation of trusted repository criteria, complemented by peer-review and other community-based assessment may be both more reliable and less burdensome that external certification. Modularization of assessment and auditing, in which audits apply to particularly subsets of criteria and responsibilities, is another promising approach. Some examples of work in this area include the *Data Seal of Approval,* which relies on peer-review of self-assessments, and the NSA Standards Group project on self-assessment and audit.

Moreover, the current trusted repository approach relies upon a very small subset of mechanisms employed in trust engineering. The role of many other approaches are both under-employed and poorly understood. *Moreover, reliability, effectiveness and costs of current trust frameworks, including TRAC and ISO16362 has yet to be empirically demonstrated and systemict*

[29] Altman, et. al "Reflections on National Digital Stewardship Alliance Member Approaches to Preservation Storage Technologies" 2013, D-Lib.

measured: How reliable are certification procedures, self-evaluations and the like at identifying good practices? How much do the implementations of such practices actually reduce risk of loss? The evidence base is not yet rich enough to answer these questions.

For example, transparency is another key mechanism for mitigation of the risks above, but it is currently underutilized and poorly understood within this domain. Implementation transparency implies the use of open protocols, and often implies the use of open source (or protocols and algorithms with independent, open implementations). Operational transparency involves demonstrating both the process and the evidence that enables others to independently verify that services are being met. The NDSA principles are an example of a generally stated organizational transparency goal, although these fall short of being enforceable.

It is particularly important that high-level policies such as TRAC can be demonstrated at the level of operations and systems action. Work such as project Pledge, the SafeArchive system, and IRods have demonstrated that complex policies can be successfully mapped to systems behavior and transparently audited.[30]

In general, further research is needed in the design, implementation and evaluation of trustworthy digital stewardship mechanisms and their use,[31] including: building an organization's capacity to demonstrate trustworthiness, rewards and penalties; peer-review; statistical quality control and reliability estimation; incentive compatible mechanisms; threat-modeling and vulnerability assessment; portfolio diversification models; transparency and the release of information permitting direct evaluation of compliance; cryptographic approaches, including cryptographic signatures over semantic content; and generating and managing social evidence of compliance.

Preservation at Scale

Digital collections are growing exponentially. Keeping track of everything and being able to work with and manage content is increasingly difficult. Growing volumes of digital materials will test the financial and operational capabilities of organizations engaged in preservation activities. Of particular concern are issues around the stewardship of big data and the search and indexing of digital collections at scale.

Institutional responsibilities to serve and preserve big data will also be influenced by user and content creator expectations regarding its maintenance and accessibility. Storage, intellectual

[30] Altman & Crabtree 2012. Smith, MacKenzie, and Reagan W. Moore. "Digital archive policies and trusted digital repositories." (2007).

[31] See B. Schneier, 2012. Liars and Outliers, John Wiley & Sons for a review of trust engineering approaches.

and administrative control, and access will all be redefined by the demands of big data. Currently, many organizations lack the expertise or economies-of-scale to store and process petabytes of data. This lack of infrastructure and expertise will require collaborative solutions involving greater automation, scalable processes, and modular, adaptive frameworks. Community-driven scalable solutions to a wide-range of unique and independent preservation activities must be developed. In addition, the establishment of shared infrastructure and open-source solutions will enable greater efficiency and economic feasibility towards the growing volume of digital content that must be preserved. Moreover, research is needed into how the stewardship of "big data" impacts privacy, confidentiality, and personally-identifiable information.

There are persistent issues in terms of indexing and searching across large amounts of content, especially while ensuring moderate reads on content stored on magnetic tape. It is no longer enough to rely on increasingly expensive and fast drives and systems. At this point, there are opportunities to exploit efficiencies in the design of smarter systems and architectures. For example, one might rebuild parts of an index when an error occurs to avoid having to restage the full index.

In sum, there is a need for collaborations with other groups and initiatives in fields addressing issues of scale in digital data, to work on common use cases and to optimize opportunities for building or acquiring cost-effective common solutions.

Strengthening the Evidence Base for Digital Preservation

A common challenge running through this report is the limited amount of empirical evidence available. For example, this report makes clear how effective digital preservation often requires answering questions such as: What is the expected future value of a specified collection of digital content? What content is already being effectively stewarded by other organizations? How much is the expected future cost of preserving that content? How often do different threats to information manifest: For example, what is the likelihood that storage hardware or media fails; software errors cause information loss; stored information becomes inaccessible because of obsolete formats, or loss of other contextual knowledge; or that human error or maliciousness causes loss content in an information system? What is the reliability of current digital preservation networks and services? And how successful are other proposed strategies for replication, monitoring, certification, and auditing at preventing loss due to these threats?

The digital preservation community is beginning to develop a shared evidence base that can be used to answer these and similar questions. Recent medium-scale observational studies and

field experiments have provided useful insights into the failure rates of spinning disk storage,[32] the proportion of files formats in use at a number of selected major digital repositories,[33] the long-term costs of preserving journal articles in pdf format,[34] and the extent and types of content being stewarded in institutional repositories.[35]

However, these studies must be broadened and repeated over time to establish a robust evidence base from which generalizable guidance can be drawn. Furthermore, for most questions in digital preservation, the current evidence base is constituted almost entirely of case studies. While case studies are useful for existence proofs, raising awareness of problems, process tracing, hypothesis generation, and other formative analysis, they are generally insufficient to advance our scientific knowledge, create robust predictive models, test causal hypotheses, or to strongly guide decision making.

For example, the NDIIPP program's highly informative case-study/field experiment in the controlled transfer of complex collections of content demonstrated the challenges of content transfer and the likelihood of failures even in well-controlled cases.[36] However, to systemically guide decisions in this area, such case studies must be repeated longitudinally, repeated in different environments, and transformed, eventually into testbeds and conformance tests that can be used to rigorously compare approaches and systems.

Similarly, bit-level preservation, which is often characterized as one of the simpler, better understood areas of preservation, lacks systematic metrics and measurements for even simple failure scenarios.[37] Furthermore there is very little information on failures in complex systems

[32] Pinheiro, E., Weber, W.D., & Barroso, L. A. (2007). Failure trends in a large disk drive population. In Proceedings of 5th USENIX Conference on File and Storage Technologies.

[33] Hitchcock, Steve, and David Tarrant. "Characterising and preserving digital repositories: File format profiles." Ariadne 66 (2011).

[34] Davies, Richard, et al. "How much does it cost? The LIFE Project-Costing Models for Digital Curation and Preservation." Liber Quarterly 17.3/4 (2007).

[35] Lynch, Clifford A., and Joan K. Lippincott. "Institutional repository deployment in the United States as of early 2005." D-lib Magazine 11.9 (2005): 5.; McDowell, Cat. "Evaluating institutional repository deployment in American academe since early 2005: Repositories by the numbers, part 2." D-lib Magazine 13.9 (2007): 3.

[36] Shirky, Clay. "Library of Congress Archive Ingest and Handling Test (AIHT) Final Report." NDIIPP. http://www. digitalpreservation. gov/partners/aiht/high/ndiipp_aiht_final_report. pdf (accessed April 22, 2011) (2005).

[37] Rosenthal, David SH. "Bit preservation: a solved problem?." International Journal of Digital Curation 5.1 (2010): 134-148.

using various redundancy, fixity, file transformation (compression, deduplication, encryption), auditing, and repair strategies. [38]

Moreover, a search of the discipline's key reference works, bibliographies, and literature databases reveal [39] very few rigorously validated preservation methods, wide-scale empirical studies, probability-based surveys or field experiments, replicable simulation experiments, public test corpuses, testbeds[40], or recognized conformance tests[41]. Broadly, across the field of digital preservation, there is an urgent need to extend the evidence base on which preservation research and policy is founded.

Evidence is needed both to support either general selection of digital preservation practices and methods, or applications of selected digital preservation methods in a specific operational context. And while preservation research should be better informed by cognate disciplines, research in information science and computer science generally target the functioning and use of information systems, and are not focused on the questions of long-term information access, understanding, and value that are central to preservation.

What is needed is to apply the research methodologies already used in other fields that rely heavily on observation of human and system behavior. This includes methodologies such as probability-based surveys of information management practice and outcomes; replicable simulation experiments tied to theoretically grounded models of information management and risk; testbeds and test-corpuses which can be used to systematically compare new practices, tools, and methods; and field experiments, in which randomized interventions are applied and evaluated in real operational environments.

[38] Baker, M., Shah, M., Rosenthal, D. S. H., Roussopoulos, M., Maniatis, P., Giuli, T., et al. (2006). A fresh look at the reliability of long-term digital storage. In Proceedings of EuroSys2006.

[39] Borghoff, Uwe M. *Long term preservation of digital documents*. Springer, 2005.; Giaretta, David, 2011, Advanced Digital Preservation, Springer. Digital Curation Center, 2012, Curation Manual, http://www.dcc.ac.uk/resources/curation-reference-manual ; Bailey Jr, Charles W. "Digital curation bibliography: Preservation and stewardship of scholarly works." *Digital Curation Bibliography: Preservation and Stewardship of Scholarly Works* (2012). Force, Blue Ribbon Task. "Sustainable economics for a digital planet: Ensuring long-term access to digital information." *Final Report of the Blue Ribbon Task Force on Sustainable Digital Preservation and Access* (2010).

[40] A notable exception being the Planets testbed: Lindley, A., Jackson, A. N., & Aitken, B. (2010, June). A Collaborative Research Environment for Digital Preservation-the Planets Testbed. In*Enabling Technologies: Infrastructures for Collaborative Enterprises (WETICE), 2010 19th IEEE International Workshop on* (pp. 197-202). IEEE. This is useful multi-institutional tool, although it provides neither the openness nor scale needed for the discipline and community.

[41] As noted above, current certifications are based primarily on process rather than demonstration of efficacy or outcome conformance.

CONCLUSION

Effective digital stewardship is vital to maintaining the nation's cultural heritage, scientific-evidence base, and the public records necessary for understanding and evaluating government actions. The aim of the 2014 National Agenda for Digital Stewardship is to engage a broad set of expertise and perspectives across the digital preservation community in order to identify productive directions for the advancement of practice and research in digital stewardship.

The 2014 National Agenda for Digital Stewardship surfaces emerging technological trends, gaps in digital stewardship capacity, and opportunities for future work for digital preservation professionals, decision makers, and others interested in investing in the long-term management of digital content. The insights provided here into emerging technological trends, gaps in digital stewardship capacity, and opportunities for funders and decision-makers will help shape the way digital content is curated and stewarded.

ABOUT THE NDSA

Founded in 2010, the National Digital Stewardship Alliance (NDSA) is a consortium of institutions that are committed to the long-term preservation of digital information. NDSA's mission is to establish, maintain, and advance the capacity to preserve our nation's digital resources for the benefit of present and future generations. NDSA member institutions represent all sectors, and include universities, consortia, professional associations, commercial enterprises, and government agencies at the federal, state, and local level.

The NDSA comprises over 150 members. These members come from 45 states and include universities, consortia, professional societies, commercial businesses, professional associations, and government agencies at the federal, state, and local level. NDSA organizations have proven themselves committed to long-term preservation of digital information. Together, these institutions have contributed over 10,000 hours of expertise to NDSA projects.

Additional materials related to the *National Agenda* and the full document can be found at: http://www.digitalpreservation.gov/ndsa/nationalagenda/index.html

BE A PART OF THE CONVERSATION

Comments welcome at ndsa@loc.gov or #nationalagenda as well as @NDSA2

ABOUT THE AUTHORS

The joint leadership group of the NDSA authored the report and engaged in discussions to identify significant trends and challenges. The membership of the NDSA contributed markedly to these discussions. This dialog was enriched by an extensive range of resources and current research. The joint leadership group is made up of the Coordinating Committee members, the Working Group co-chairs, and the NDSA facilitator:

Micah Altman

Director of Research, MIT Libraries, MIT; Non-Resident Senior Fellow, Brookings Institution.

Dr Micah Altman is Director of Research and Head/Scientist, Program on Information Science for the MIT Libraries, at the Massachusetts Institute of Technology. Dr. Altman is also a Non-Resident Senior Fellow at The Brookings Institution. Dr. Altman conducts research in social science, information science and research methods—focusing on the intersections of information, technology, privacy, and politics, and on the dissemination, preservation, reliability and governance of scientific knowledge. Altman serves on the NSDA Coordinating Committee.

Jefferson Bailey

Strategic Initiatives Manager, Metropolitan New York Library Council

Jefferson Bailey manages programs for the the library, archive, and museum communities. He was formerly a Fellow in Digital Preservation at the Library of Congress, a manager of digital projects at Brooklyn Public Library and Frick Art Reference Library, and has done archival work at NARA, NASA, and The New York Times. He is on the Curriculum Committee of the National Digital Stewardship Residency program, co-chair of the NDSA Innovation Working Group, and is the and a visiting lecturer in University of Pittsburgh's Library & Information Science program.

Karen Cariani

Director of Media Library and Archives, WGBH

Karen Cariani has worked at WGBH since 1984 in television production and archival-related roles. She has 20-plus years of production and project management experience, has worked on numerous award-winning historical documentaries, and has been project director for many critical projects. She worked with the WNET, PBS, NYU and WGBH Preserving Public Television partnership as part of the Library of Congress National Digital Information Infrastructure Preservation Project. She served two terms (2001-2005) on the Board of Directors of Association of Moving Image Archivists (AMIA). She also serves on Digital Commonwealth executive committee. Recent projects include managing the American Archive Inventory project for CPB, and project director of PBCore development and Boston Local TV News Digital Library project. Cariani is the co-chair of the NDSA Infrastructure Working Group.

Jim Corridan

State Archivist and Director of the Indiana Commission on Public Records

Jim Corridan was elected to the Board of Directors of the Council of State Archivists (CoSA) in 2009, where he helped establish and chaired CoSA's State Electronic Records Initiative (SERI) in 2011 and currently serves as President of CoSA. SERI is focused on governance, best practices, awareness, and education to strengthen all the state and territorial archives in the United States. Working with the staff from the Library of Congress, Corridan and the Indiana State Archives hosted the first regional Digital Preservation Outreach and Education (DPOE) project, the Midwest train-the-trainer program, in August of 2012 and has been an advocate for digital preservation. He is State Archivist of Indiana.

Jonathan Crabtree

Assistant Director of Computing and Archival Research at the Odum Institute for Research in Social Science at UNC Chapel Hill

As assistant director, Jonathan Crabtree completely revamped the institute's technology infrastructure and has positioned the institute to assume a leading national role in information archiving. Crabtree's 22 years of experience in information technology and networking as well as his engineering background bring a different perspective to his current role. Crabtree serves on the NSDA Coordinating Committee.

Blane Dessy

Director, Federal Library and Information Network (FEDLINK)

Blaine Dessy's career has included appointments as a public library director, state library consultant, state librarian, Federal program officer for library grants, Director of the ERIC program, founding director of the National Library of Education, and library director at the Department of Justice. He has been involved in several professional associations including ALA, SLA, AALL, IFLA, GreyNet, the American Society for Public Administration, and the American Political Science Association. Currently, he is an adjunct instructor at the Catholic University of American School of Library Science. Mr. Dessy has lectured on librarianship in Turkey and Brazil and has represented the United States at education conferences in Italy and the Netherlands. Dessy serves on the NSDA Coordinating Committee.

Michelle Gallinger

Digital Programs Coordinator, National Digital Information Infrastructure & Preservation Program, Library of Congress

Michelle Gallinger works to develop digital preservation communities. Gallinger develops policies and guidelines for digital preservation practices, life-cycle management of digital materials, and stakeholder engagement at the Library of Congress. She also provides strategic planning for the National Digital Information Infrastructure and Preservation Program. Gallinger developed the initial strategy for and supported the creation, definition, and launch of the National Digital Stewardship Alliance in 2010.

Andrea Goethals

Digital Preservation and Repository Services Manager, Harvard Library

Andrea Goethals is responsible for providing leadership in the development and operation of Harvard's digital preservation program and for the management and oversight of the Digital Repository Service (DRS), the university's large-scale digital preservation repository. She is currently involved in the rolling out of the next-generation DRS ("DRS2"), several pilot projects (email archiving, exposing public APIs to access DRS content and metadata), and planning for new digital preservation services at the university. She is on the Curriculum Committee of the National Digital Stewardship Residency program and participates in the International Internet Preservation Consortium's Preservation Working Group. She is the co-chair of the NDSA Standards and Practices Working Group.

Abbie Grotke

Lead Information Technology Specialist, National Digital Information Infrastructure & Preservation Program, Library of Congress

Abbie Grotke is the Web Archiving Team Lead at the Library of Congress and a member of the NDIIPP team. She came to the Library in 1997 to work on American Memory digitization projects. Since 2002 she has been involved in web archiving and the digital preservation program at the Library of Congress. Grotke currently serves on the Steering Committee of the International Internet Preservation Consortium and is the co-chair of the NDSA Content Working Group.

Cathy Hartman

Associate Dean of Libraries, University of North Texas

Cathy Hartman's preservation efforts began in 1997 by establishing the CyberCemetery to preserve the Websites of U.S. government agencies and commissions and creating partnerships at Federal and State levels to archive electronic government information. She founded the Portal to Texas History to enable collaborations with more than 200 libraries and museums to digitize and preserve their collections for Web access. She currently chairs the International Internet Preservation Consortium Steering Committee and co-chairs the NDSA Content Working Group.

Butch Lazorchak

Digital Archivist, National Digital Information Infrastructure and Preservation Program, Library of Congress

Butch Lazorchak is a digital archivist in the National Digital Information Infrastructure and Preservation Program at the Library of Congress. He sits on the NDIIPP Communications Team, working to expand awareness of digital preservation issues and opportunities through a variety of channels, including regular postings for The Signal blog. He is the co-chair of the NDSA Outreach Working Group.

Jane Mandelbaum

Manager of Special Projects, Office of the Director for Information Technology Services, Library of Congress

Jane Mandelbaum is currently leading and guiding enterprise-wide projects and architecture initiatives for large-scale, high-performance digital storage and archiving. She previously served as IT implementation and operations manager for a number of large IT systems at LC and led a team to establish and operate the Library's end-user computing environment. Mandelbaum is the co-chair of the NDSA Innovation Working Group.

Carol Minton Morris

Director of Marketing and Communications, DuraSpace

Minton Morris (Terrizzi) joined the National Science Digital Library (NSDL) team at Cornell University in 2000 and served as Communications Director from 2000--2009. She was the Communications Director for the Fedora Commons organization from 2007--2009. As Director of Marketing and Communications for the DuraSpace organization since 2009 she leads strategic editorial content and materials planning, development and distribution focused on sustaining open source projects (DSpace, Fedora) and marketing services (DuraCloud, DSpaceDirect). She is deputy co--chair of the annual International Open Repositories Conference and serves as co--chair of the NDSA Outreach Working Group.

Trevor Owens

Digital Archivist, National Digital Information Infrastructure & Preservation Program, Library of Congress

At the Library of Congress, Trevor Owens works on the open source Viewshare cultural heritage collection visualization tool, as a member of the communications team, and on convening relevant stakeholders about preservation issues. Before joining the Library of Congress he worked for the Center for History and New Media and before that managed outreach for the Games, Learning, and Society Conference. Owens is the co-chair of the NDSA Infrastructure Working Group.

Meg Phillips

Electronic Records Lifecycle Coordinator, National Archives and Records Administration

Meg Phillips assists senior management with Electronic Records Archives (ERA) planning and the coordination of electronic records units and projects. In addition, Phillips is responsible for gathering and helping NARA prioritize the business requirements for the development and evolution of ERA. Phillips serves on the NSDA Coordinating Committee.

John Spencer

President and co-founder, BMS/Chace.

John Spencer has widespread experience and visibility both in the music industry and in the fields of archival preservation and enterprise class information technology. Since 1978, he has been involved in many facets of high-technology professional audio and video, and was previously Vice President of Sales and Marketing for Otari Corp. Spencer is CEO of BMS/Chace, working with commercial companies to further the value of structured metadata collection for media companies and institutions worldwide. He is a member of the following professional associations: The Recording Academy Producers & Engineers Wing National Advisory Council, and Nashville chapter Advisory Council, Audio Engineering Society (AES) Studio Practices and Production Technical Committee, Association of Recorded Sound Collections (ARSC) Technical Committee, National Recording Preservation Board (NRPB) Digital Audio Preservation, and Standards Task Force. Spencer serves on the NSDA Coordinating Committee.

Helen Tibbo

Alumni Distinguished Professor, School of Information and Library Science (SILS) at the University of North Carolina at Chapel Hill (UNC-CH).Professor Helen Tibbo teaches in the areas of archives and records management, digital curation, electronic retrieval, and reference. She is an SAA fellow, has served on SAA committees and boards for over 20 year, was the co-founder of the SAA Research Forum, and was SAA President, 2010-2011. She is currently the primary investigator for the IMLS-funded CRADLE project that is creating learning tools for archivists, librarians, and science researchers regarding data management. She was PI for DigCCurr I and II, which developed Digital Curation curriculums for master's level students (www.ils.unc.edu/digccurr) (2006-2009), doctoral students and information professionals (2008-2013). Tibbo serves on the NSDA Coordinating Committee.

Tyler Walters

Dean of University Libraries, Virginia Tech.

Tyler Walters was a 2008-2010 Fellow in the Association of Research Libraries' Research Libraries Leadership Fellows program. Walters is a founding Board member of the Educopia Institute and Steering Committee member of the MetaArchive Cooperative. He serves on many professional bodies, such as the Board of Directors for DuraSpace, Board of Directors for the National Information Standards Organization (NISO), Steering Committee of the Coalition of Networked Information (CNI) and Editorial Board of the International Journal of Digital Curation. Walters has presented at numerous conferences, has published over twenty-five journal articles, and is a recipient of the Society of American Archivists' Ernst Posner Award for best article in the American Archivist. Walters serves on the NSDA Coordinating Committee.

Kate Wittenberg

Managing Director, Portico

Before taking on the leadership of Portico, Kate Wittenberg was Project Director, Client and Partnership Development in Ithaka S+R, where she focused on building partnerships among scholars, publishers, libraries, technology providers, and scholarly societies with an interest in promoting the development and sustainability of digital scholarship and learning. Before joining Ithaka, she directed the Electronic Publishing Initiative at Columbia, a collaboration of the libraries, academic computing, and university press.. Kate serves on the NDSA Coordinating Committee.

Kate Zwaard

Supervisory Information Technology Specialist, Repository Development Center, Library of Congress

Kate Zwaard manages software development projects in the Repository Development Center, which creates applications to support the digital content lifecycle. Prior to joining LC, she was the Lead Program Planner for digital preservation at the United States Government Printing Office, where she helped conceptualize and develop the Federal Digital System. She has served on the PREMIS editorial committee. Zwaard co-chairs the NDSA Standards and Practices Working Group